ON GENERALIZATION OF WIENER INDEX FOR CYCLIC STRUCTURES

Milan Randić

National Institute of Chemistry, Ljubljana, Slovenia; e-mail: milan.randic@ki.si and Department of Mathematics and Computer Science, Drake University, Des Moines, IA 50311, USA Home FAX: 515 292 8629

Received 15-11-2001

Abstract

We have critically examined a particular generalization of Wiener index to cyclic structures known as Szeged index. Limited ability of Szeged index to give fair correlation of boiling points of cycloalkanes was traced to a deficiency in its definition which leads to great variation in index values between odd member cyclic systems and even member cyclic systems, both of which show similar boiling points. A remedy is suggested for construction of Wiener index more suited for cyclic structures.

Mathematical elegance is not enough for good topological index.

Introduction

Mathematical descriptors of molecular structure, such as various topological indices.¹ have been widely used in structure-property-activity studies. This includes the multiple regression analysis (MRA), the principal component analysis (PCA),² the pattern recognition,³ the artificial neural networks (ANN),⁴ optimization of lead compound and combinatorial optimization of the lead compound,⁵ search for pharmacophore,^{6,7} and the similarity-dissimilarity studies. Apparently there is no lack of initiative and imagination in proposing novel topological indices, but there appears to be lack of efforts to demonstrate of their specificity and lack of efforts in clarification of their structural meaning. By specificity we understand here useful characterization of one of molecular properties by the novel index when used alone, or in combination with other molecular descriptors. By structural meaning of an index we understand interpretation of an index in terms of elementary structural concepts such as atoms, Only recently^{8,9} such an bonds, or in terms of paths and walks of longer length. interpretation was offered for several well known indices, including the connectivity index,¹⁰ the Wiener index,¹¹ and the Hosoya Z topological index.¹² Lists of desirable

qualities for novel topological indices have been proposed,^{13,14} but have also been mostly ignored or overlooked. In this article we will focus attention on generalization of the Wiener index for use in cyclic structures. As is well known, Wiener¹¹ defined his W index only for acyclic graphs. Attempts have been reported to generalize Wiener index to cyclic graphs. An index proposed by Gutman,¹⁵ known as Szeged index, received considerable attention.¹⁶⁻²³ It has attractive mathematical characteristics, but does it offer *useful* characterization of cyclic structures?

In this paper we have critically re-examined the Szeged index and found it deficient as a molecular descriptor for structure-property-activity studies, despite it's elegant and attractive mathematical definition. We will present its "structural" deficiency and will show how the observed deficiency can be corrected. This resulted in a novel index, revised Wiener index RW, which as will be seen shows better descriptor for structure – property relationship for cyclic molecules.

One may say that there are essentially two reasons for considering novel or alternative molecular descriptors: (1) Novel descriptors may lead to better regression analysis, and (2) Novel or alternative descriptors may be computationally simple, which is important when screening combinatorial libraries which may have 100,000 structures or more. To this we would like to add an additional consideration: (3) Novel or alternative descriptors may offer a simpler structural interpretation. There are hundreds of available mathematical descriptors for molecular graphs, many of which can be computed by programs such as MOLCONN,²⁴ POLLY,²⁵ CODESSA.²⁶ For a recent compilation of large number of topological indices readers are directed to a book of Todeschini and Consinni.²⁷ In Table 1 we have collected several recently introduced topological indices still based on rather simple structural components, merely to illustrate structural diversity of topological indices. Among those listed is also the revised Szeged index, and the revised "higher order" Szeged indices (analogous to the hyper-Wiener index of acyclic structures).

In this paper we will closely examine the revised Szeged index, but as will be apparent similar modifications considered here for Szeged index apply also to the "higher order" indices which use paths of different length instead or considering only bonds (paths of length one). Finally, we should add that most of distance related indices can be associated with matrices, which are either sparse (when considering only adjacent atoms) or dense (when considering paths of different length). A number of novel graph matrices have been introduced in recent years that include for instance the Wiener matrix,²⁸ the Hosoya Z matrix,²⁹ the Restricted Random Walk matrix,³⁰ the Distance/Distance matrix,³¹ the Resistance-Distance matrix,³² the Detour matrix,³³ two kinds of Path matrices,^{34,35} etc. In this way recently a dozen novel atomic descriptors were generated. We should add that the first such matrices beyond the traditional adjacency and distance matrix are the Expanded Wiener matrix of Tratch, Stankevich and Zefirov,³⁶ and the electrotopological matrix of Kier and Hall.³⁷

Table 1. Several novel topological indices based on use of different structural factors

index	structural factors involved
Shape index w/p	given by quotient of paths and walks of a same length
Double invariants	matrix elements given by subgraphs; numerical values by subgraph invariants
Composite X'/X descriptors	index is given by quotient of an invariant X and the same invariant when one bond at a time is excluded from graph
Variable connectivity	Parameters x, y, z,, which characterize different kinds of atoms, are optimized during regression process
Chirality index	Derived from invariants obtained by circumscribing planar molecules in opposite direction
Generalized (acyclic) Wiener	Multiply the number of atoms on each side of a path for every pair of vertices
Generalized (cyclic) Wiener	Multiply the number of atoms on each side of an edge closer to each end atoms, add 1/2 of atoms at equal distance
Generalized (cyclic) Wiener	Multiply the number of atoms on each side of a path closer to each end atoms for every pair of vertices, add 1/2 of atoms at equal distance

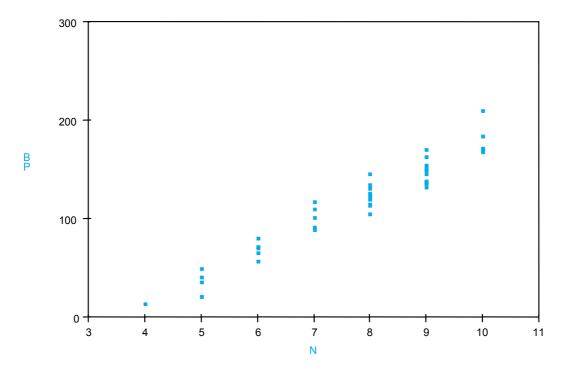
Szeged index

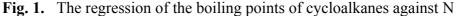
Wiener,¹¹ has defined his W index as a bond additive index in which each bond makes a contribution $(m \times n)$, where m and n are the number of atoms on each side of a bond considered. This definition allows a generalization^{28,38,39} in which all pair of atoms makes contribution, not only adjacent atoms. One simply replaces "(m x n), where m and n are the number of atoms on each side of a bond considered" by " $(m \times n)$, where m and n are the number of atoms on each side of a path considered." In this way a matrix is obtained, the so-called Wiener matrix,38 from which the hyper-Wiener index^{38,39} was extracted. This generalization, however, has not considered cyclic structures, which constitute a large portion of chemistry. Gutman has considered generalization of W to cyclic structures and came with mathematically elegant definition of the Wiener index for cyclic structures that bears some analogy to already mentioned design of hyper-Wiener index. Gutman has replaced "(m x n), where m and n are the number of atoms on each side of a bond considered" in the definition of Wiener by "(m x n), where m and n are the number of atoms closer to atoms on each side of a bond considered." Clearly in cyclic atoms one can not speak of "of atoms on each side of a bond" because bond does not separate cyclic structure in two parts. However, in both cases, acyclic and cyclic compounds, one can speak of "of atoms closer to atoms on each side of a bond."

What is wrong with Szeged index

Szeged index has an elegant mathematical formulation but from its limited applications it appears not to offer useful correlation. First observe that of some dozen papers on Szeged index most of these are concerned with mathematical properties of this index, its relationship to other indices, and its evaluation for special classes of molecules rather then demonstrating its use in chemical applications. Very few of these contributions mention any correlation with any of physicochemical properties examined by other indices. It is nothing wrong with studying mathematical properties of graph invariants, but unless it first is demonstrated that a particular invariant has some use in the structure-property-activity relationships such mathematical studies are premature, to say the least.

The only paper that we found in which Szeged index is tested in QSPR (quantitative structure-property relationship) is paper by Diudea⁴⁰ in which he examined use of Szeged index (and several other distance-based indices) on correlation with the boiling points (BP) of 45 cycloalkanes. When in a simple regression one uses N, the number of carbon atoms in a molecule, one obtains regression shown in Fig. 1 characterized by the regression coefficient r = 0.969, the standard error s = 10.9; and the Fisher ratio F = 664.





Clearly all isomers (having the same N) are grouped together displaying limitations of N as molecular descriptor to differentiate isomeric variations. Any index that does not do better than N, which is only a measure of molecular size, can be viewed, at least for the particular application, as useless. The simple regression of the boiling points of cyclic alkanes based on the Szeged index and shown in Fig. 2 is characterized by the following statistical parameters: r = 0.919, s = 17.7; and F = 225 - hence, useless, being worse than the regression based on N, the number of carbon atoms in a structure.

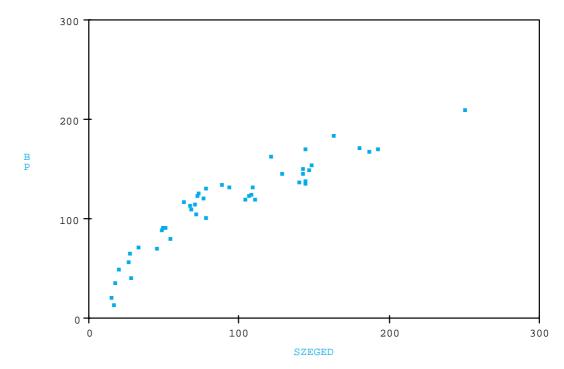


Fig. 2 Regression of the boiling points of cycloalkanes against Szeged index

isomer	BP	Szeged	isomer	BP	Szeged
Even rings			Odd rings		
11MC6	119.5	104	112MC5	114	67
12MC6	123.4	106	113MC5	105	71
13MC6	124.5	108	123MC5	115	70
14MC6	120	110	12MEC5	124	72
EC6	131.8	109	13MEC5	121	76
C8	146	128	PC5	131	78
			IPC5	126.4	73
			MC7	134	88
Average	127.5	110.8	Average	121.3	74.4
Average*	123.8	107.4	Average*	119.5	72.4

Table 2. The boiling points and Szeged indices for N = 8 cycloalkanes isomers

Why is that Wiener index (in combination with polarity index P) has produced outstanding regressions for several physico-chemical properties of alkanes and other compounds as described over 50 years ago by Wiener while its generalization to cyclic compounds is not producing good correlation at all? The problem is, as we have found, with the proposed generalization of the Wiener index as offered by Gutman. In Table 2, we have listed for cycloalkanes having N = 8 their boiling points and their Sz indices separating compounds having even and those having odd member rings. As we see from Table 2 on average the BP of cycloalkanes having even member ring and those having odd member ring do not differ greatly. On the other hand the average the Szeged numbers for the two kind of cycloalkanes show an excessive difference in their relative magnitude. This is even more pronounced in the last row of Table 2 in which we have confined attention only to five and six member ring compounds. The average BP of the two groups differ by about 4 °C (about 3%), while the Szeged index of the two groups differ by 35 units (more than 30%). Clearly the Szeged index cannot be successive in correlating properties which do not depend dramatically on the parity of ring size.

Improved Wiener index for cyclic structures

Now that we pointed to inadequacy of Szeged index we will outline a design of an improved or *revised* Wiener index (RW) for cyclic structures. From Table 2 it can be inferred that problem is with the Szeged index for compounds having *odd rings*, the magnitude of which are too small, and not with the Szeged index of compounds with even member rings. This is based on the fact that Wiener index for acyclic compounds having the same number of CC bonds (e. g. n-nonane) is around 100 and not around 75. One can immediately recognize that reduced size of Szeged index of compounds having odd rings is due to neglect of contributions from atoms (and bonds) which are at the same distance from both atoms forming a bond. Hence, we have to augment Szeged index with contributions from atoms not considered in the definition of Sz index as proposed by Gutman. A simple remedy to deficiency of Sz is to divide equally the count atoms at the same distance from atoms at both ends of a bond. This is illustrated in Fig. 3 on 1-methyl-2-ethylcyclopentane (12MEC5). As we see, instead of having

Sz = 72 we obtain RW = 100, which is close to Szeged indices for isomers having even member ring. In this way we obtain the revised Szeged indices of Table 3 for cycloalkanes having odd ring.

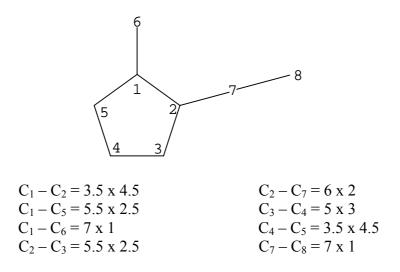


Fig. 3 Calculation of revised Wiener index RW for 1-methyl-2-ethylcyclopentane

M = methyl; E = ethyl; P = propyl; IP = isopropyl					
	Szeged	RW		Szeged	RW
11MC3	15	24.75	113MC5	71	99
EC3	17	26.75	123MC5	70	98.5
C5	20	31.25	12MEC5	72	100
112MC3	26	40.5	13MEC5	76	104
123MC5	27	42	PC5	78	105
MC5	33	49	IPC5	73	100
11MC5	48	69.25	MC7	88	117.5
12MC5	49	70.75	1123MC5	93	128.75
13MC5	51	72.75	EC7	121	157.75
C7	63	85.75	С9	144	182.25
112MC5	67	95	PC7	163	207.5

Table 3. Szeged index and the Revised Wiener index RW for cyclic structures M = methyl: F = ethyl: P = propyl: IP = isopropyl

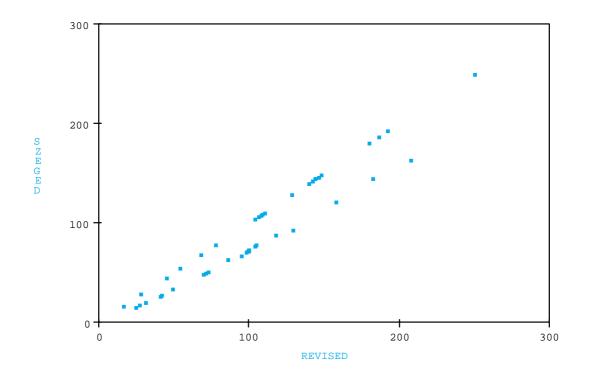


Fig. 4 Correlation between Szeged index and Revised Wiener index for cyclic structures

In Fig. 4 we show a correlation between the Szeged index as defined by Gutman and others and the revised Szeged index RW for the 45 cycloalkanes considered by Diudea in his paper on Cluj matrices. From the figure one immediately sees a "double" line pattern, one given by equation y = x, and the other below it showing some minor departures from a strict line. The upper line belongs to cyclic alkanes having even ring, in which case the Szeged index and RW are identical. Compounds having even member ring need no correction because in such systems always an atom is closer to one end of any CC bond. The small oscillations in the lower line are caused by different number of "ignored" atoms in computation of the Szeged index. For example, for 1MC7 only one carbon (of the methyl group) has been "ignored" in calculating the contribution for a CC bond on the opposite site of methyl group in the 7-member ring. In contrast in PMC5 four carbon atoms have been "ignored" in calculating the contribution for a CC bond on the opposite site of the propyl group in 5-member ring.

Correlation of BP for cycloalkanes with Revised Wiener index RW

In Fig. 5 we show correlation between BP of cyclkoalkanes and RW, the revised Wiener index for cyclic structures.

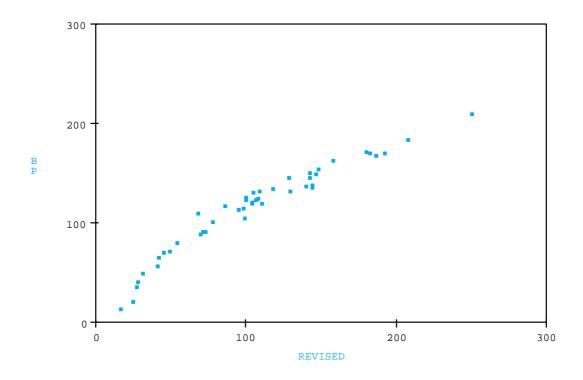


Fig. 5 Regression of the boiling points of cycloalkanes against RW index

A comparison of Fig. 5 with Fig. 2 immediately shows that regression based on the revised Wiener index RW is much better. In Table 4 we have summarized statistical information on the regression using Sz and RW. As we see a quadratic regression gives for the standard error 8.3 °C as compared with 14.3 °C. This seems an impressive improvement of the regression. While this may still not be satisfactory regression from the point of view of structure-property correlation as a simple regression based on a single descriptor it certainly represents a satisfactory preliminary result. For example, recently⁴¹ a comparison is made between a correlation of the boiling points of cycloalkanes using the connectivity index¹⁰ and the variable connectivity index.⁴²⁻⁴⁹ In both cases the standard error was close to 3 °C , showing that bond additive descriptors can offer satisfactory correlation with boiling points in the case of cycloalkanes.

Descriptor	r	S	F	
N	0.9667	11.4	614	
Szeged index	0.9207	17.4	239	Linear
	0.9480	14.3	186	Quadratic
Revised	0.9635	11.9	556	Linear
Wiener				
	0.9828	8.3	596	Quadratic

Table 4. The statistical data for regression of BP of 45 cycloalkanes using the Sz and RW. Included is statistical data for regression using N (illustrated in Fig. 1)

The Wiener index and the Szeged index, as well as here considered their modifications, even though can be viewed also as bond additive molecular descriptors, have totally different structural origin, being based on distance metrics. It is thus of considerable interest to see how these distance-related indices behave. Szeged index, which represents a particular generalization of the Wiener index to cyclic structures, failed to produce even rough correlation. However, this apparently has not been either recognized or reported in the literature to warn potential users. Thus we here want to draw attention of those interested in topological indices that clearly "something is rotten in the state of Denmark," that is in "proposed extention of the Wiener index to cyclic structures." While there is no doubt that the Szeged index has mathematical elegance, mathematical elegance alone is not enough for an index to qualify as useful topological index!

	BP calc	Residual		BP calc	Residual
C4	28.5	-15.4	11MC6	120.8	-1.3
11MC3	39.3	-18.3	12MC6	122.5	0.9
EC3	41.7	-5.8	13MC6	124.1	0.3
MC4	43.2	-2.6	14MC6	125.8	-5.8
C5	47.0	2.3	EC6	125.0	6.8
112MC3	57.8	-1.3	MC7	131.8	2.2
123MC3	59.5	6.5	C8	139.8	6.2
EC4	62.6	7.8	1123MC5	140.3	-7.6
MC5	67.3	4.5	113MC6	148.2	-11.6
C6	72.8	7.9	124MC6	150.9	-14.9
PC4	87.4	22.6	135MC6	150.9	-12.4
11MC5	88.6	0.3	12MEC6	149.6	1.4
12MC5	90.1	1.8	13MEC6	152.2	-3.2
13MC5	92.1	-0.4	PC6	153.5	0.5
MC6	97.3	3.6	IPC6	149.6	-3.6
C7	104.6	12.4	EC7	159.5	4.0
112MC5	113.0	1.0	C9	172.7	-2.7
113MC5	116.5	-11.5	12MIPC6	171.6	-0.6
123MC5	116.1	-1.1	13MIPC6	174.5	-7.0
12MEC5	117.4	6.6	13EC6	177.2	-6.7
13MEC5	120.8	0.2	PC7	183.4	0.1
PC5	121.6	9.4	C10	194.7	15.3
IPC5	117.4	9.0			

Table 5. Computed BP and associated residuals for 45 cycloalkanes using RW indices

For a completeness of the presentation in Table 5 we show computed BP and associated residuals for 45 cycloalkanes using RW indices. As we can see four cycloalkanes are associated with the largest positive and the largest negative residuals. When these four compounds (#1, #2, #11, and #45) are removed we obtain respectable single variable regression characterized by the regression coefficient r = 0.9877, the standard error s = 6.0 °C, and Fisher ratio F = 760. We should mention that regression using Sz index has much larger scatter of points that it does not allow one to identify potential outliers, thus one is stuck with the standard error of 14.3 °C, much too large to be of use in discussion of relative boiling points of cycloalkanes.

Concluding Remarks

Now that we pointed to inadequacy of Szeged index and have outlined a design of Revised Wiener index for cyclic structures we hope that researches in the field will recognize the importance of "use" of topological indices and the study of their limitations, rather than merely "studying their mathematical properties," which can only be of interest if topological index has found use. In the past there has been too much emphasis in some circles on novelty of indices, as if a novelty is guarantee of usefulness of such indices in chemical applications. The present extension of the Wiener index to cyclic structures applies equally to construction of matrices associated with distances in cyclic systems. In a follow up paper⁵⁰ we have re-examined extension of here presented generalization of the Wiener index to the higher order Wiener indices based on considerations of paths of longer length and atoms at equal distance from such paths. This similarly leads to a revision of previously proposed indices in which atoms and bonds at equal distance from the end points of paths have been ignored, in our view incorrectly ignored.

References and Notes

- 1. M. Randic, Topological Indices. *Encyclopedia of Computational Chemistry* (P. von Rague Schleyer, Editor-in-Chief, London: Wiley, 1998, pp.3018-3032.
- 2. H. Hotelling J. Educ. Psychol. 1933, 24, 8417 & 489.
- 3. T. Okuyama, Y. Miyashita, S. Kanaya, H. Katsumi, S. I. Sasaki, and M. Randic J. Comput. Chem. 1988, 9, 636.
- 4. J. Zupan, J. Gasteiger, Neural Networks for Chemists; VCH: Weinheim, 1993.
- G. Grassy, B. Calas, A. Yasri, R. Lahana, J. Woo, S. Iyer, M. Kaczorek, R. Floc'h, R. Buelov Nature Biotechnology, 1997, 70, 819.
- 6. M. Randic Molecular Basis of Cancer, Part A: Macromolecular Structure, Carcinogens, and Oncogens, (R. Rein, Ed.), Alan Liss Publ., 1985, pp. 309-318.
- 7. M. Randic, B. Jerman-Blazic, D. H. Rouvray, P. G. Seybold, S. C. Grossman Int. J. Quantum Chem: Quantum Biol. Symp. 1987, 14, 245.
- 8. M. Randic, J. Zupan J. Chem. Inf. Comput. Sci. 2001, 41, 550.
- 9. M. Randic, J. Zupan, On Structural Interpretation of Topological Indices, in: *Proc. Wiener Memorial Conference*, Athens 2001 (R. B King and D. H. Rouvray, Eds.)
- 10. M. Randic J. Am. Chem. Soc. 1975, 97, 6609.
- 11. H. Wiener J. Am. Chem. Soc. 1947, 69, 17.
- 12. H. Hosoya Bull. Chem. Soc. Jpn. 1971, 44, 2332.
- 13. M. Randic J.Math. Chem. 1991, 7, 155.
- 14. A. T. Balaban J. Mol. Struct. (Theochem), 1988, 165, 243.
- 15. I. Gutman Graph Theory Notes, 1994, 27, 9.
- 16. I. Lukovits, W. Linert J. Chem. Inf. Comput. Sci. 1994, 34, 899.
- 17. P. V. Khadikar, N. V. Deshpande, P. P. Kale, A. A. Dobrynin, I. Gutman, G. Domotor J. Chem. Inf. Comput. Sci. 1995, 35, 547.
- 18. I. Gutman S. Klavzar J. Chem. Inf. Comput. Sci. 1995, 35, 1011.
- 19. D. J. Klein. I. Lukovits, I. Gutman J. Chem. Inf. Comput. Sci. 1995, 35, 50.

M. Randić: On generalization of Wiener index for cyclic structures

- 20. M. V. Diudea, O. Minailiuc, G. Katona, I. Gutman MATCH, 1997, 35, 000.
- 21. J. Zerovnik Croat. Chem. Acta, 1996, 69, 837.
- 22. A. A. Dobrynin Croat. Chem. Acta, 1997, 70, 819.
- 23. I. Lukovits J. Chem. Inf. Comput. Sci. 1998, 38, 125.
- 24. L. H. Hall MOLCONN-X; Hall Associates Consulting, Quincy: MA, 1991
- 25. Basak, S. C. *POLY*; Natural Resources Research Institute, Duluth, University of Minnesota: Duluth, MN, 1988.
- 26. Katritzky, A. R.; Lobanov, V.; Karelson, M. CODESSA (Comprehensive Descriptors for Structural and Statistical Analysis); University of Florida: Gainesville, FL, 1994.
- R. Todeschini, V. Consonni Handbook of Molecular Descriptors (Methods and Principles in Medicinal Chemistry, vol. 11, (R. Mannhold, H. Kubinyi, H. Timmerman, Eds.), Wiley-VCH.
- 28. M. Randic Chem. Phys. Lett. 1993, 221, 478.
- 29. M. Randic Croat. Chem. Acta 1994, 67, 415.
- 30. M. Randic Theor. Chim. Acta 1995, 92, 97.
- 31. M. Randic, A. F. Kleiner, L. M. DeAlba J. Chem. Inf. Comput. Sci. 1994, 34, 277.
- 32. D. J. Klein, M. Randic J. Math. Chem. 1993, 12, 81.
- 33. D. Amic, N. Trinajstic Croat. Chem. Acta 1995, 68, 53.
- 34. M. Randic, D. Plavsic, M. Razinger MATCH 1997, 35, 243.
- 35. O. Ivanciuc, A. T. Balaban MATCH 1994, 30, 141.
- 36. S. S. Tratch, M. I. Stankevitch, N. S. Zefirov J. Comput. Chem. 1990, 11, 899.
- 37. L. B. Kier, L. H. Hall Molecular Structure Description: The Electrotopological State: Academic Press (1999).
- 38. M. Randic, X. Guo, T. Oxley, H. Krishnapriyan J. Chem. Inf. Comput. Chem. 1993, 33, 709.
- 39. M. Randic, X. Guo, T. Oxley, H. Krishnapriyan, L. Naylor J. Chem. Inf. Comput. Chem. 1994, 34, 361.
- 40. M. Diudea J. Chem. Inf. Comput. Sci. 1997, 37, 300.
- 41. M. Randic, D. Plavsic, N. Lers J. Chem. Inf. Comput. Sci. 2001, 41, 657.
- 42. M. Randic Chemom. Intell. Lab. Syst. 1991, 10, 213
- 43. M. Randic J. Comput. Chem. 1991, 12, 970.
- 44. M. Randic, J. Cz, Dobrowolski Int. J. Quantum Chem: Quantum Biol. Symp. 1998, 70, 1209.
- 45. M. Randic, D. Mills, S. C. Basak Int. J. Quantum Che:. 2000, 80, 1199.
- 46. M. Randic, S. C. Basak J. Chem. Inf. Comput. Sci. 2000, 40, 899.
- 47. M. Randic, S. C. Basak J. Chem. Inf. Comput. Sci. 2001, 41, 619.
- 48. M. Randic, M. Pompe J. Chem. Inf. Comput. Sci. 2001, 41, 631.
- 49. M. Randic New J. Chem. 2000, 24, 165.
- 50. M. Randic, M. Novic, M. Vracko, J. Zupan J. Chem. Inf. Comput. Sci. (to be submitted)

Povzetek

Kritično smo pregledali posebno posplošitev Wienerjevega indeksa za ciklične strukture, imenovano Szeged index. Ugotovili smo, da je izvor omejene zmožnosti Szegedovega indexa, da bi izkazal dobro korelacijo z vrelišči cikloalkanov, v njegovi pomanjkljivi definiciji. Posledica take definicije je velika variacija vrednosti indeksa, izračunanega za ciklične strukture, katerih obroči imajo liho oziroma sodo število atomov, njihova vrelišča pa so podobna. Predlagamo popravek, ki nudi bolj primerno rešitev za gradnjo Wienerjevega indeksa za ciklične strukture.